

Initial MAC Exploration for Graphene-enabled Wireless Networks-on-Chip

G. Piro¹, S. Abadal², A. Mestres², E. Alarcón², J. Solé-Pareta², L. A. Grieco¹, G. Boggia¹

¹ DEI, Politecnico di Bari, Via Orabona 4, 70124, Bari, Italy

² Departament d'Arquitectura de Computadors Campus Nord - UPC, Jordi Girona 1-3, 08034 Barcelona, Spain

ABSTRACT

In the upcoming *many-core era*, chip multiprocessor architectures will be composed by hundreds or even thousands of processor cores, which interact among them through an on-chip communication platform for synchronization and data coherency/consistency purposes. As the traffic generated within the chip becomes more multicast-intensive, it is necessary to conceive novel communication platforms that go beyond conventional schemes and guarantee multicast support with high throughput, low latency, and low power. Nanotechnology provides an opportunity within this context by virtue of terahertz graphene antennas, which could allow the integration of one antenna per core in a Graphene-enabled Wireless Network-on-Chip (GWNoC). However, it is essential to design an appropriate MAC protocol in order to fully benefit from this novel approach. To provide a first contribution in this direction, in this paper we design two baseline MAC protocols based on the well-known ALOHA and carrier sensing techniques. Their functionalities have been properly conceived by taking into account characteristics and requirements of future chip multiprocessors systems. Moreover, their performances have been evaluated by means of computer simulations under different chip configurations. Obtained results demonstrate the pros and cons of these simple contention-based MAC protocols and pave the way for the future exploration of the MAC design space.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous; D.2.8 [Software Engineering]: Metrics—*complexity measures, performance measures*

Keywords

WNoC, MAC protocol, Nanonetworks, Performance Evaluation

1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM NANOCOM, Atlanta, Georgia, USA, May 13-14 2014
Copyright 2014 ACM.

Diminishing returns in instruction-level parallelism have recently caused the emergence of multiprocessor architectures [1]. A Chip Multiprocessor (CMP) system represents a particular case of such trend and consists in the interconnection of multiple independent processor cores within a single chip. The *on-chip interconnect* is a central element within this context since it implements the communication between cores and memory and has a large impact on performance. Consequently, significant research efforts have been devoted to the design of scalable and efficient solutions for these interconnects. Buses were firstly considered in literature. However, their adoption is restricted to small-scale architectures because of the limited scalability beyond a few cores. Nowadays, instead, Network-on-Chip (NoC) represents the dominant paradigm for on-chip interconnects [2, 3]. It can be defined as the application of the principles of packet switching networking to on-chip communications.

In the upcoming *many-core era*, where processors will be composed of hundreds or even thousands of cores, new challenges will emerge in the NoC design process related to both the limitations in chip area/power and the sharply increase of communication requirements [4]. One particularly concerning issue is related to the significant growth in point-to-multipoint communications within CMPs. Conventional NoCs do not efficiently support this type of traffic, since they generally convert a multicast message into multiple and parallel unicast transmissions, resulting in significant penalties in terms of performance and power consumption [5]. Seeking to avoid on-chip communication in general (and point-to-multipoint communication in particular) to become a bottleneck in CMPs, current research efforts in this area are focused upon finding new interconnect technologies that would either complement or supplant traditional NoCs to deliver the required performance.

The Wireless Network-on-Chip (WNoC) approach, which enables wireless communications among cores, represents one important breakthrough in this direction as it guarantees high flexibility, low latency, and low power consumption in chip-wide transmissions, as well as a native support to *one-to-many* communications [6, 7]. Unfortunately, current WNoC proposals cannot fully benefit from the native broadcast capabilities of the approach since size constraints preclude the possibility of integrating one antenna per core (see Sec. 2 for more details).

The WNoC approach could benefit from constant advances in nanotechnology, which are enabling the design and development of wireless communication units at the nanoscale. This would indeed pave the way for the novel and very

promising on-chip communication paradigm known as Graphene Wireless Network-on-Chip (GWNOC) [6]. GWNOC is expected to resolve the aforementioned fundamental issue in the pathway to efficiently satisfying the challenging requirements of future CMPs by virtue of graphene antennas [6]. Unique plasmonic effects cause a 5 μm long and 1 μm wide graphene antenna to radiate in the terahertz band [8], both enabling broadcast capabilities at the core-level through the integration of one antenna per core, and providing data rates up to tens of Terabits per second (Tbps) [9].

The ambitious goal set by GWNOC requires the design of a Medium Access Control (MAC) protocol capable of properly handling the simultaneous exchange of messages among cores. This represents a great challenge since, first, the medium is shared by hundreds or even thousands of communication-intensive cores. Moreover, the propagation of electromagnetic waves in the terahertz channel implies a large propagation-to-transmission ratio, thus generating new open issues related to the design of efficient channel access procedures. As a consequence, the medium access control cannot be resolved by applying existing MAC mechanisms for WNoC, which generally consist in a combination of time and frequency multiplexing schemes [7]. Contention-based MAC mechanisms could be used instead, given their higher flexibility and improved scalability in terms of complexity.

To the best of the authors' knowledge, contention-based mechanisms have not been explored within the GWNOC context. To bridge this gap, this work aims to set the foundations on this emerging and challenging research activity. Starting from the analysis of characteristics and requirements of future CMP architectures, we highlight the main design criteria that should be carefully taken into account to create the communication layer of a GWNOC system. Then, we conceive two simple MAC protocols based on the well-known Aloha and Carrier Sense Multiple Access (CSMA) techniques, which regulate the contention among cores in a chip. Finally, we evaluate the performance of the devised solutions in different chip configurations by means of simulation. We used NANO-SIM, an open-source tool modeling electromagnetic-based communications at the nanoscale [10], to carry out the simulations.

Obtained results, will demonstrate the pros and cons that the investigated architecture provides in different cases. In particular, we verified that while high performance is achievable assuming a limited traffic load, simple contention-based approaches is not well-suited to very high loaded scenarios. For this reason, our findings open the research community to new and exciting challenges related to the design of sophisticated MAC strategies for future massive multicore systems.

The rest of this paper is organized as follows. Section 2 briefly reviews related work about on-chip networking, MAC protocols for WNoC, and simulators for communications at the nanoscale. The conceived communication layer, including two MAC protocols, is described in Section 3. A performance evaluation of conceived solutions is shown and discussed in Section 4. Finally, Section 5 concludes the paper and outlines future research.

2. BACKGROUND AND RELATED WORKS

2.1 Multicast in NoC and WNoC

A CMP is a system consisting of a number of cores (i.e., the computing layer), a memory system (i.e., the I/O layer),

and an interconnection network (i.e., network layer). The main current trend in the network layer design is based upon the NoC paradigm, which consists of a network of electrical on-chip wires and simple routers that form a given topology. Conventional NoCs are typically based upon point-to-point links and, as a consequence, are not suited to multicast/broadcast communications. Broadcast transmissions are in fact treated by generating copies of the same message and conveying them to the intended receivers.

With the advent of *many-core* architectures, point-to-multipoint communications within CMPs are expected to experience a sharp increment. In this context, conventional NoCs will yield a performance degradation both in terms of power consumption and computational speed [5]. To alleviate such effects, some works suggest to implement a set of virtual multicast trees within the network layer of a CMP [5, 11, 12, 13]. Even though these proposals are able to reduce power requirements and latency for multicast transmissions, their scalability remains largely unexplored.

As highlighted above, the emerging WNoC paradigm provides, among other advantages, a potential platform for the efficient service of broadcast transmissions. This is mainly due to its shared medium nature, which, at the same time, limits the usefulness of the approach in the presence of high loads of unicast messages. In light of this, it is reasonable to assume that a WNoC will be juxtaposed to a conventional NoC as depicted in Figure 1. In such hybrid network architecture, each core will be directly connected to both the wireless and wired planes and will account for a controller that decides through which plane a message needs to be sent.

In order to guarantee the delivery of simultaneous broadcast transmissions in WNoC, a MAC protocol is required. Very simple solutions presented in literature manage medium access by means of channelization techniques, which are based on Frequency Division Multiple Access (FDMA) [14, 15], Time Division Multiple Access (TDMA) [16, 17], and Code Division Multiple Access (CDMA) [18]. The main downturn of these approaches is the limited scalability given by the strong trade-off between transceiver complexity and number of channels. Systems that do not rely on channelization schemes have been discussed in [19] and [20]. In the former contribution, a *token ring* scheme is used to synchronize the access to the channel. Only the node that has the *token* is able to transmit. The token is passed to the next node either when the node has nothing else to transmit or the time slot finishes. Since the token passes through all nodes, such scheme presents limited scalability. In the latter contribution, a contention-based channel access protocol is described. In that scheme, each core announces an imminent transmission by sending a specific message through a wired NoC, thus reducing the probability to generate wireless collisions. However, the work considers a transmission range limited to a reduced set of neighbors and the protocol is therefore not applicable to a general setting. Moreover, such approach requires multi-hop to reach furthest nodes, therefore increasing the latency of the transmissions.

2.2 Towards the GWNOC Paradigm

The main limitation of the WNoC approach refers to the impossibility of integrating at least one antenna within each core, as future metallic antennas will be hundreds of micrometers long [15] and cores continue to shrink until reaching sizes of a few hundreds of micrometers. Although it is pos-

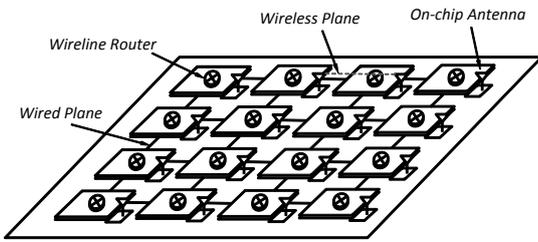


Figure 1: Hybrid wired-wireless architecture.

sible to scale WNoC architectures for many-core CMPs by dividing the network into clusters and by using wireless and wired interfaces for inter- and intra-cluster communications, respectively [16, 21], the final architecture will experience limited performances in terms of latency and power consumptions. Antenna size is, therefore, the main barrier that will prevent WNoC systems to provide core-level broadcast capabilities in future massive multicore systems.

As nanotechnology will enable the manufacturing of micrometer graphene antennas, the foundations for the emerging Graphene Wireless Network-on-Chip (GWNOC) paradigm are set. Differently from in a WNoC, it is possible to deploy one antenna within each core thanks to the reduced size of the graphene antennas. In addition, these antennas radiate in the terahertz band, guaranteeing a sufficient capacity for massive multicore settings [9]. Intense research efforts have been recently directed towards accurately modeling these antennas in order to obtain their characteristics in terms of antenna impedance, bandwidth, radiation efficiency and radiation pattern [22, 23].

Fig. 2 shows a GWNOC architecture, where each core is equipped with a graphene antenna and a wireless transceiver [6] (a wired NoC is also deployed but not shown for the sake of clarity). Graphene antennas are composed of a finite-size graphene layer, which is mounted over a metallic flat surface (the ground plane) by means of a dielectric material, and an ohmic contact. The transceiver is in charge of preparing the information for outgoing transmissions and demodulates incoming transmissions. As remarked in [24], the communication in the terahertz channel can be executed using Impulse Radio (IR) techniques. By exchanging very short pulses (i.e., each one lasting some femto or pico seconds), signals are spread over the entire THz spectrum.

The use of graphene antennas within the GWNOC context also impacts upon the definition and expected performance of the medium access control protocol. For instance, the extremely high transmission speeds that could be enabled by graphene antennas imply a potentially large propagation-to-transmission ratio, fact that will drastically influence the contention mechanisms inside the chip. To the best of the authors' knowledge, this is the first work discussing the application of MAC protocols in the GWNOC scenario.

2.3 Nanoscale Communication Simulators

Nowadays, researchers worldwide are exploring novel protocol stacks, network architectures, and channel access procedures that could be adopted at the nanoscale. In this context, a number of simulation tools could support these activities. Most of them, i.e., NanoNS [25], N3Sim [26], and the one proposed in [27], have been explicitly conceived for diffusion-based molecular communications. Indeed, NANO-

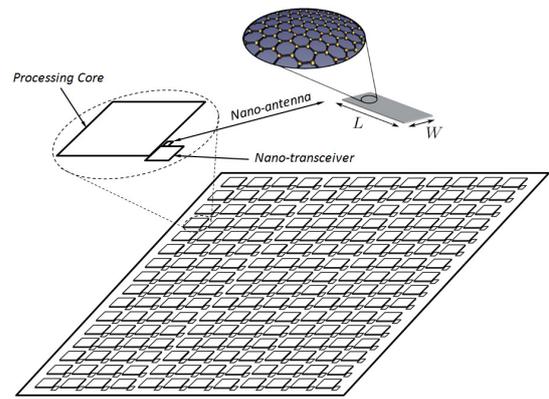


Figure 2: Chip Multiprocessor architecture based on the Graphene Wireless Network-on-Chip paradigm [6].

SIM represents the only available simulator modeling EM-based nanonetworks [28][29]. For this reason, it is the only research tool currently available to study GWNOC architectures.

NANO-SIM has been developed on top of NS-3, a discrete-event and open-source network simulator, designed with the aim of replacing the popular NS-2 in both research and educational fields. The source code of NANO-SIM is freely available under the GPLv2 license, thus boosting its diffusion in the research community [10]. NANO-SIM was initially conceived to simulate wireless nanosensor networks (WNSNs). In particular, at the time of this writing, it implements (i) different kinds of devices forming a WNSN, (ii) a physical interface based on the Time Spread On-Off Keying (TS-OOK) modulation scheme, (iii) two different MAC protocols, namely *Transparent-MAC* and *Smart-MAC*, (iv) a routing module handling both selective flooding and random strategies, and (v) a generic unit for generating and processing messages. Recently, the IEEE P1906.1 WG [30], focusing on nanoscale and molecular communications, has identified NANO-SIM as one of reference simulation platform for electromagnetic communication at the nanoscale. Thanks to its high flexibility, NANO-SIM can be properly upgraded to model nano-communications also in the GWNOC context.

3. PROPOSED COMMUNICATION LAYER FOR GWNOC ARCHITECTURES

The GWNOC architecture we conceived in this work should efficiently support the execution of parallel instructions in a CMP. Given that processes executed in parallel need to coordinate their access to specific remote memory locations, the network layer is mainly used to:

- during the execution of a process, communicate any modifications on shared variables (data coherency issue);
- if a given variable is shared among different processes, ensure the correct modification order (data consistency issue);
- synchronized before continuing with the execution since, in many algorithms, the programmer requires that all

processing cores wait for a given line of code to be executed (synchronization issue).

These issues generate the exchange of a huge amount of messages within the chip and the characteristics of this traffic will depend on the particular implemented CMP architecture. Without loss of generality, we can assume that conventional architectures generate a heterogeneous mix of unicast and broadcast messages that must be served by the communication layer with a high level of reliability.

Given that the GWNOC paradigm is more suited for multicast communications due to its inherent shared medium nature, and in line with the architecture envisaged in [6], we designed a communication layer composed by two interfaces: the *wireless plane* and the *wired plane*. The former is used to handle broadcast transmissions, whereas the latter is based on a conventional NoC and deals with unicast transmissions.

Communication is carried out as follows. When a core has a message to transmit, its controller checks whether the message is unicast or broadcast and delivers it to the corresponding interface. In the wired plane, messages are simply routed to their destination. In the wireless plane, the wireless interface broadcasts the corresponding packet within the chip. As destination cores receive the packet, they generate an ACK of confirmation that will be delivered to the sender through the wired plane. The source collects and counts the ACKs. If any ACK is missing, the source node determines that the message was not received by all cores, probably due to physical errors or collisions). In such case, the source node will decide, depending on the retransmission policy, to broadcast again the packet through the wireless interface or to deliver the message only to cores that have not received it by means of wired transmissions.

To control the access to the shared channel in the wireless plane, we conceived two different MAC protocols: *ALOHA-based* and *CSMA-based*. As their names imply, they are based on the well-known ALOHA and CSMA techniques, respectively. These strategies has been designed assuming that the maximum wired delivering delay (W_D) is approximately the same for both one-to-all (broadcast) and all-to-one (ACK) flows [31, 32].

Table 1 summarizes the set of parameters that will be adopted in the sequel.

3.1 ALOHA-based Protocol

Based on the pure ALOHA protocol, this represents one of the simplest MAC strategies that can be used in a GWNOC. As a consequence, its performance may be considered a lower bound because of the very limited resilience of the protocol to wireless collisions.

In the *ALOHA-based* protocol, a message received from upper layers is immediately sent through the wireless plane, without previously checking the status of the channel. The core that handled the transmission will collect ACKs generated by receiving cores during a specific time interval, namely timeout, i.e., T_O . If it is verified that all the destination cores have correctly received the transmitted message, the transmission will be considered completed and the sender can definitively delete the packet from the MAC queue. Otherwise, at the expiration of the timeout, the message will be transmitted through the wired plane, either again to all the cores or to only those cores that have not received it in the past. It is worth noting that we disabled

Table 1: Adopted parameters

Parameter	Description
W_D	maximum delay achievable in the wired plane
T_X	transmission time, i.e., the time required to transmit a data packet by using the wireless interface
T_O	timeout value, which triggers a retransmission procedure
N_r	number of retransmissions executed through the wireless interface
N_r^{max}	maximum number of retransmissions allowed for the wireless interface
T_P	maximum propagation delay in the wireless plane
M_{Bt}	maximum backoff time
B_T	backoff time computed by the MAC protocol
r	random number used for computing the backoff time

any kind of wireless retransmission in order to ensure that the MAC protocol fully respects, for the *wireless plane*, the normal behavior of the well-known ALOHA protocol. The reliability of the communication is guaranteed by the packet retransmission handled through the *wired plane*.

To avoid useless wired transmissions, it is important to correctly size the timeout interval. In order to take into account the delays introduced by both the wireless and wired planes, we computed the T_O value as follows:

$$T_O = W_D + T_P + T_X. \quad (1)$$

3.2 CSMA-based Protocol

In the *CSMA-based* protocol, the well known CSMA strategy is used to before transmitting a message in the wireless plane. Also, the *CSMA-based* protocol adopts an advanced retransmission procedure and a backoff strategy in order to improve the overall performance of the GWNOC. Differently from the simple *ALOHA-based* approach, the packet retransmission is allowed through the wireless interface. In particular, the core tries to deliver the message through the *wireless plane* provided that the maximum number of wireless retransmissions, N_r^{max} , is not reached. In the case this threshold is exceeded and the packet transmission is still incomplete, the message is delivered through the *wired plane*.

Before sending a packet through the wireless interface, the source core senses the channel in order to identify the presence of other active communications. On the one hand, if the channel is perceived in an idle state, the wireless transmission is executed. Similarly to in the *ALOHA-based* scheme, the sender will collect ACKs of confirmation until all ACKs are received or upon the expiration of the timeout (see Eq. (1)). On the other hand, if channel is busy, the core will wait of a backoff time before starting again the channel sensing procedure. The backoff time, B_T , is computed as in the following:

$$B_T = r \cdot M_{Bt}, \quad (2)$$

where r is a random number between $[0,1]$ and M_{Bt} is eval-

uated considering the number of retransmissions performed for that particular packet, N_r , and the maximum wireless propagation delay, i.e.,

$$M_{Bt} = 2^{N_r} \cdot T_P. \quad (3)$$

In the case the transmission has not been completed at the expiration of the timeout (i.e., ACKs have not been received from all the destination cores), a wireless retransmission will be scheduled after a new backoff time. To this end, the core increases by one unit the N_r variable, evaluates the new M_{Bt} value, and computes the new backoff time interval as reported in Eq. 2. It is important to note that, in order to limit the delay caused by the backoff strategy in congested situations, the protocol considers a maximum number of allowed wireless retransmissions. As soon N_r exceeds it (i.e., $N_r > N_r^{max}$), the core will complete the transmission through the wired plane, like the *ALOHA-based* scheme.

If the transmission of a given message is completed by using the wireless interface, the transmission procedure of a new packet is triggered immediately after the reception of the last ACK from the previous transmission. Otherwise, the consecutive transmission will be scheduled after a backoff time, which is computed with Eq. 2 and by setting $N_r = 0$.

4. PERFORMANCE EVALUATION

Performances of both *Aloha-based* and *CSMA-based* protocols have been evaluated through computer simulations, by using the emerging NANO-SIM simulator [28].

4.1 Extensions Implemented in NANO-SIM

To implement the protocols proposed in this paper, NANO-SIM has been extended in three parts: the application layer, the communication interface, and the MAC entity.

In our analysis we assumed that: (i) all the broadcast packets exchanged within the chip have the same size; (ii) their generation rate is constant, and (iii) generation time instants are independent on each other. To model these characteristics, a Poisson-based source has been implemented at the application layer. Two parameters are used to characterize its behavior: the packet size, i.e., p_s (which is expressed in bits), and the average source rate, i.e., s_r (which is expressed in bps). Indeed, according to the Poisson distribution, the inter-arrival time is modeled through an exponential random variable with parameter λ :

$$\lambda = \frac{s_r}{p_s}. \quad (4)$$

In NANO-SIM, a device is conceived as a container of a set of entities, such as the message processing unit, the routing layer, the MAC layer, and the physical interface. Thus far, it was only possible to connect devices to a channel that is in charge of delivering packets through electromagnetic waves in the terahertz band. The communication interface was therefore composed by only one physical layer and the wireless channel, which interacted between them to perform transmission and reception procedures [29]. In line with the GWNOC communication paradigm described in Sec. 2.2, we extended the communication interface of NANO-SIM. Now, the simultaneous presence of *wireless* and *wired planes* is possible.

On the one hand, the physical interface of the *wireless plane* has been modeled considering an On-Off Keying (OOK) modulation scheme. The choice is driven by its simplicity

[24]. According to the simulation model presented in [29], the channel handles the physical transmission considering the time instant when the transmission starts, the transmission duration (which depends of both the pulse duration of the OOK modulation scheme and the packet length) and the propagation delay (which is computed considering the distance among cores and the propagation speed of the light). This way, the time instant in which the message is delivered to a remote wireless interface can be calculated, and possible packet collisions in the wireless channel can be identified.

On the other hand, we assumed that all the cores are connected among through a conventional meshed NoC in the *wired plane*. We modeled this plane from a system level point of view and focused on the wireless plane. To this end, we just introduced the W_D variable for defining the maximum delivering delay available for the wired channel. In this preliminary study, in fact, we will consider W_D as a load-independent parameter, thus postponing the introduction of load-dependent models for specific acknowledging schemes with diverse performance ranges in future works.

Finally, two new MAC entities have been developed to model the conceived channel access procedures.

4.2 Network Configurations and Simulation Parameters

By taking into account the current state-of-the-art of CMP systems (e.g. TILERA commercial products are made by up to 64-72 cores [33]) and the typical scaling trends that assume an increment of the number of cores between 1.4X and 2X for each technology generation [34], we considered in our study a multiprocessor composed of a number of cores, i.e., from 64 to 576, uniformly distributed on a chip of area 20 by 20 mm².

Cores run an application that generates synthetic broadcast traffic at a source rate between 0.1 and 10 Gbps for each core. Such rates are easily achievable in memory-intensive applications and assuming broadcast-based coherency and synchronization methods. Moreover, two packet size values have been considered in our analysis. i.e., 312 bits and 1024 bits, seeking to model both short control messages and long data messages.

Communication between cores is performed by means of the hybrid wired-wireless architecture suggested in Section 2. By using the PhoenixSim framework [35], we verified that latencies in the wired plane may range from tens to a hundreds of nanoseconds, depending from the number of cores and the traffic load. In line with these findings, we evaluated the presented architecture by setting W_D to 10 and 100 nanoseconds.

With respect to the wireless plane, we assumed a GWNOC where all cores are in the same transmission range. At the physical layer, we used an OOK modulation scheme with a pulse duration set to 10^{-12} s. Such duration is justified by the expected frequency of radiation of the considered antennas, and leads to a potential maximum throughput of 1 Tbps. At the MAC layer, we consider either the *ALOHA-based* or the *CSMA-based* protocol, setting the maximum number of allowed wireless retransmissions to 2 in the latter case.

Furthermore, simulation results have been averaged over 10 consecutive runs, thus reducing the impact of statistical fluctuations.

To conclude, we note that the propagation-to-transmission

ratio, which is determined by the chip size, the packet size and the maximum throughput, will yield a value between 0.1 and 0.3 for this configuration. In conventional wireless communication scenarios, it is widely proved that CSMA-like protocols outperform the ALOHA protocol when the propagation time is shorter than the transmission time. Therefore, it is expected that the *CSMA-based* protocol will outperform the *ALOHA-based* one in the chip communication scenario.

4.3 Analysis of Results

The percentage of transmissions completed by using only the wireless plane represents the first important performance metric we considered in our study. As reported in Fig. 3, we found that it decreases as both the number of cores and the average source rate increase. As expected, more loaded scenarios imply a higher congestion level in the wireless channel, with a consequent impairment of the overall performance. Moreover, it is observed that the higher is the packet size, the larger is the amount of transmissions completed by using only wireless communications. In this case, the core transmits a lower number of messages for a fixed average source rate, thus reducing the probability to incur in wireless collisions. More importantly, we found that the *CSMA-based* approach enables an improved use of the wireless channel thanks to the implemented retransmission procedure and the backoff strategy. This demonstrates, as expected, that the *CSMA-based* protocol better exploits the available wireless bandwidth and reduces the amount of broadcast communications performed through the wired plane. Since the wireless plane is potentially more efficient than the wired plane for broadcast transmissions, this result suggests that the *CSMA-based* protocol could lead to a higher reduction of the overall latency and power consumption than the *ALOHA-based* protocol. It is important to note that the W_D parameter has a very limited impact upon the percentage of successful wireless transmissions, especially at lower loads. Finally, and despite all the aforesaid advantages, the simple *CSMA-based* protocol is not able to guarantee very good performances in chips with a high number of cores or with high traffic loads.

Figure 4 reports the aggregate network goodput obtained by the wireless plane in all the considered network configurations. This metric is closely related to the number of successful wireless transmissions and the packet size. In line with previous results, the *CSMA-based* proposal guarantees a higher performance thanks to its capability to better exploit the available bandwidth at the wireless plane. The *ALOHA-based* protocol shows an evident degradation of its performance after a given *saturation* load, whereas sensing techniques enables the *CSMA-based* option to maintain a constant goodput in saturation. As discussed above, larger packets result in slightly higher goodput in all cases due to the reduced number of collisions.

Finally, we have also evaluated the average packet delay (see Figure 5). The delay is computed as the difference between the time instant in which the transmission has been completed and the packet generation time. It is important to remind that the Acknowledgment (ACK) messages are delivered through the wired plane. Additionally, a broadcast packet will be transmitted through the wired plane in the presence of collisions: the *ALOHA-based* protocol relies on the wired plane directly after the first timeout, whereas the *CSMA-based* protocol tries to use the wireless plane twice

again before relying upon the wired plane. For all this, the performance of the wired plane (modeled by means of the W_D parameter) has a large impact upon the delay. For a low W_D , the *ALOHA-based* approach outperforms the *CSMA-based* protocol for large aggregated loads. In this case, it is worth noting that we considered a load-independent W_D ; however, transmitting broadcast messages through the wired plane would create additional contention in the wired network and affect the performance of concurrent transmissions (increase W_D). For a high W_D , it is observed that the *CSMA-based* protocol reduces the average delay in all cases.

5. CONCLUSIONS

In this paper, we studied one key aspect in the adoption of the GWNOC paradigm within massive multicore architectures: the MAC protocol. Assuming a hybrid wired-wireless network architecture, we conceived two baseline MAC protocols based on the well-known ALOHA and CSMA techniques. The performance of these solutions was evaluated through computer simulations considering different network sizes, traffic intensities, packet sizes and capabilities of the wired plane. From the analysis of obtained results, we verified that the *CSMA-based* protocol outperforms the simple *ALOHA-based* approach, because of its capability to better exploit the wireless channel when the propagation-to-transmission ratio is lower than one. Since the wireless plane is potentially more efficient than the wired plane for broadcast transmissions, this result suggests that the *CSMA-based* protocol could lead to a higher reduction of the overall latency and power consumption than the *ALOHA-based* protocol. Despite these very interesting advantages, we observe that a simple *CSMA-based* protocol is not able to guarantee very good performances in chips with a high number of cores or with high traffic loads. For this reason, in the future we will investigate alternative channel access strategies for the emerging GWNOC paradigm, aiming to further support the requirements of future massive multicore architectures.

6. ACKNOWLEDGMENTS

This work has been partially supported by the FI-AGAUR grant of the Catalan Government, by the PON projects (RES NOVAE, ERMES-01-03113, DSS-01-02499 and EURO6-01-02238) funded by the Italian MIUR and by the European Union (European Social Fund), as well as by INTEL through its Doctoral Student Honor Program.

7. REFERENCES

- [1] J. Hennessy and D. Patterson, *Computer architecture: a quantitative approach*. 2012.
- [2] W. Dally and B. Towles, "Route packets, not wires: on-chip interconnection networks," in *Proceedings of the 38th IEEE Design Automation Conference*, pp. 684–689, Acm, 2001.
- [3] L. Benini and G. De Micheli, "Networks on chips: a new SoC paradigm," *Computer*, vol. 35, no. 1, pp. 70–78, 2002.
- [4] J. Owens, W. Dally, R. Ho, D. Jayasimha, S. Keckler, and L. Peh, "Research challenges for on-chip interconnection networks," *Micro, IEEE*, vol. 27, no. 5, pp. 96–108, 2007.
- [5] N. E. Jerger, L.-S. Peh, and M. Lipasti, "Virtual Circuit Tree Multicasting: A Case for On-Chip

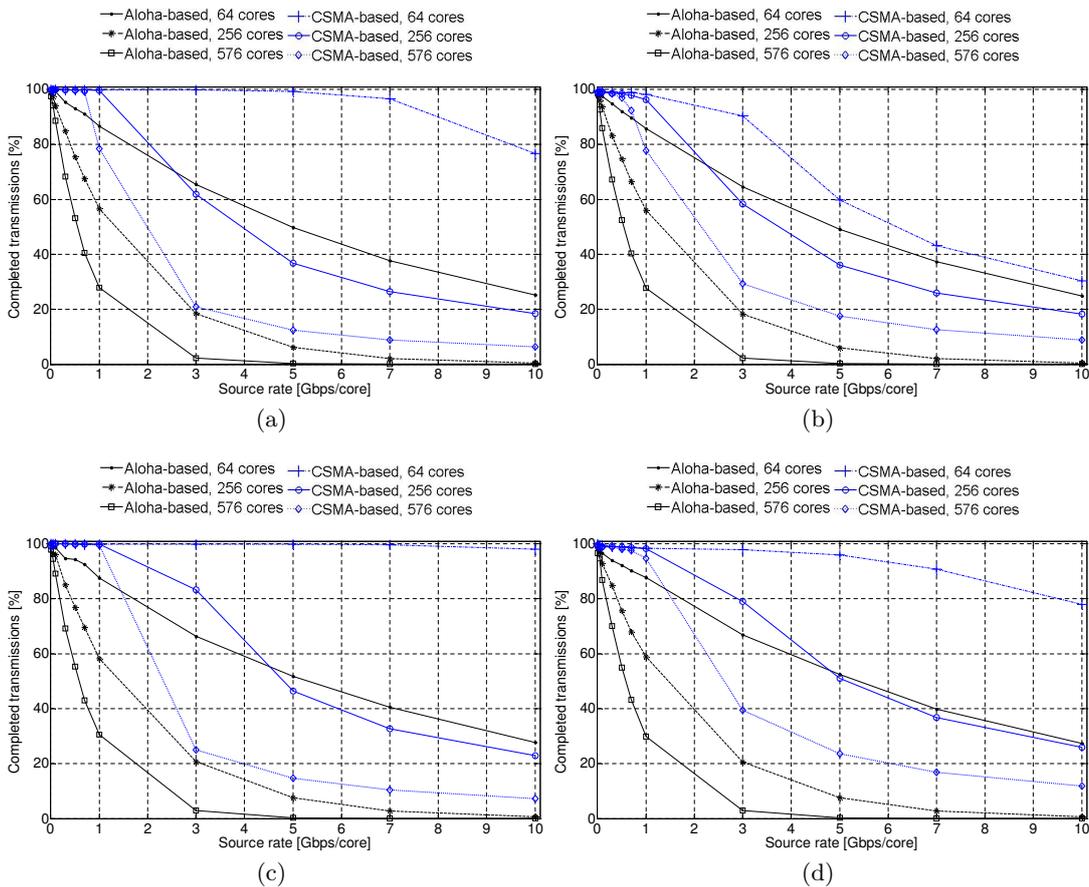


Figure 3: Percentage of successful wireless transmissions, evaluated when (a) packet size = 312 bits and $W_D = 10$ ns, (b) packet size = 312 bits and $W_D = 100$ ns, (c) packet size = 1024 bits and $W_D = 10$ ns, and (d) packet size = 1024 bits and $W_D = 100$ ns.

Hardware Multicast Support,” *2008 International Symposium on Computer Architecture*, pp. 229–240, June 2008.

- [6] S. Abadal, E. Alarcón, M. C. Lemme, M. Nemirovsky, and A. Cabellos-Aparicio, “Graphene-enabled Wireless Communication for Massive Multicore Architectures,” *IEEE Communications Magazine*, vol. 51, no. 11, pp. 137–143, 2013.
- [7] S. Deb, A. Ganguly, P. P. Pande, B. Belzer, and D. Heo, “Wireless NoC as Interconnection Backbone for Multicore Chips : Promises and Challenges,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems (JETCAS)*, vol. 2, no. 2, pp. 228–239, 2012.
- [8] J. M. Jornet and I. F. Akyildiz, “Graphene-based nano-antennas for electromagnetic nanocommunications in the terahertz band,” in *Proc. of 4th European Conference on Antennas and Propagation (EUCAP, ed.)*, (Barcelona), 2010.
- [9] J. M. Jornet and I. F. Akyildiz, “Channel Modeling and Capacity Analysis for Electromagnetic Wireless Nanonetworks in the Terahertz Band,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3211–3221, 2011.
- [10] G. Piro, “Nano-sim - The open source framework for

simulating EM-based WNSNs.” [OnLine] Available: <http://telematics.poliba.it/nano-sim>.

- [11] S. Rodrigo, J. Flich, J. Duato, and M. Hummel, “Efficient unicast and multicast support for CMPs,” *2008 41st IEEE/ACM International Symposium on Microarchitecture*, pp. 364–375, Nov. 2008.
- [12] F. A. Samman, T. Hollstein, and M. Glesner, “Multicast parallel pipeline router architecture for network-on-chip,” in *Proceedings of the Conference on Design, Automation and Test in Europe (DATE)*, pp. 1396–1401, ACM Press, 2008.
- [13] L. Wang, Y. Jin, H. Kim, and E. Kim, “Recursive partitioning multicast: A bandwidth-efficient routing for Networks-on-Chip,” in *International Symposium on Networks-on-Chip (NoCs)*, pp. 64–73, 2009.
- [14] E. Tavakoli, M. Tabandeh, S. Kaffash, and B. Raahemi, “Multi-hop communications on wireless network-on-chip using optimized phased-array antennas,” *Computers & Electrical Engineering*, vol. 39, pp. 2068–2085, July 2013.
- [15] S.-B. Lee, L. Zhang, J. Cong, S.-W. Tam, I. Pefkianakis, S. Lu, M. F. Chang, C. Guo, G. Reinman, C. Peng, and M. Naik, “A scalable micro wireless interconnect structure for CMPs,” *Proceedings of the 15th annual international conference on Mobile*

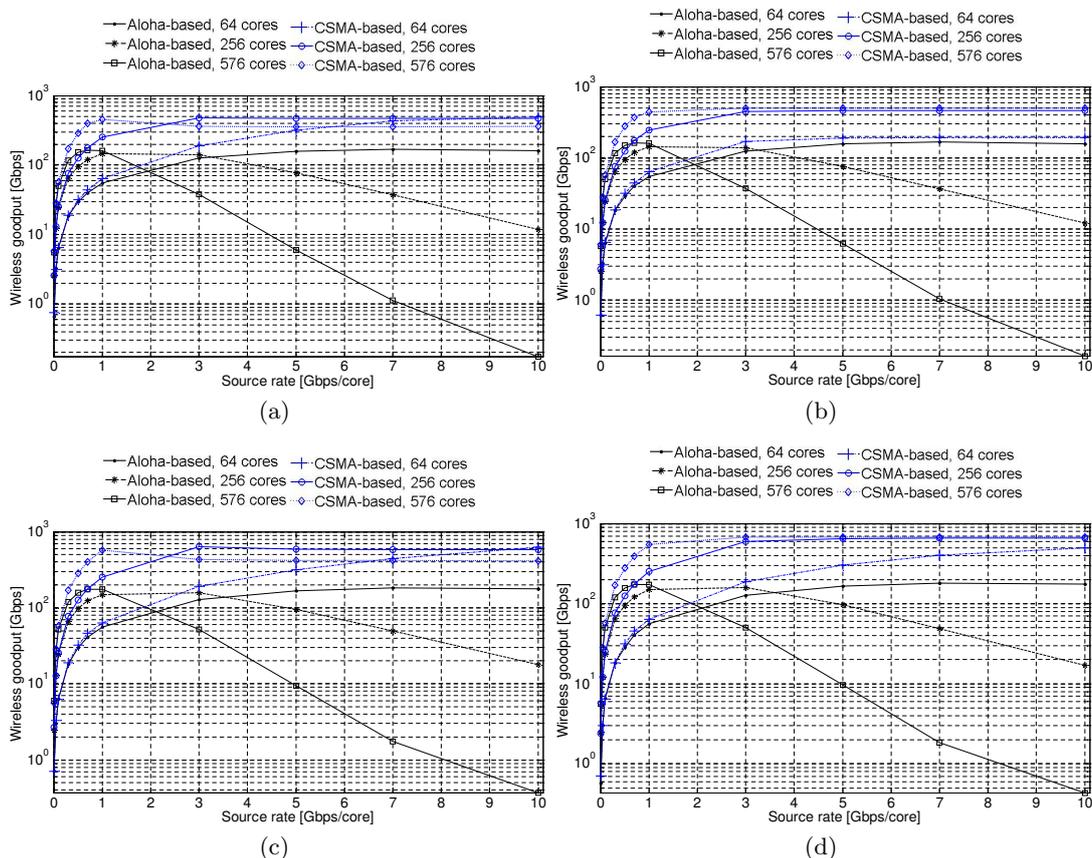


Figure 4: Wireless goodput, evaluated when (a) packet size = 312 bits and $W_D = 10$ ns, (b) packet size = 312 bits and $W_D = 100$ ns, (c) packet size = 1024 bits and $W_D = 10$ ns, and (d) packet size = 1024 bits and $W_D = 100$ ns.

computing and networking - MobiCom '09, p. 217, 2009.

- [16] D. W. Matolak, A. Kodi, S. Kaya, D. DiTomaso, S. Laha, and W. Rayess, "Wireless Networks-on-Chip: Architecture, Wireless Channel, and Devices," *Wireless Nanoscale Communications*, no. 5, pp. 58–65, 2012.
- [17] A. Ganguly, S. Deb, and B. Belzer, "Scalable hybrid wireless network-on-chip architectures for multicore systems," *Computers, IEEE Transactions on*, vol. 60, no. 10, pp. 1485–1502, 2011.
- [18] A. Vidapalapati, V. Vijayakumaran, A. Ganguly, and A. Kwasinski, "NoC architectures with adaptive Code Division Multiple Access based wireless links," *2012 IEEE International Symposium on Circuits and Systems*, pp. 636–639, May 2012.
- [19] S. Deb, A. Ganguly, K. Chang, P. Pande, B. Beizer, and D. Heo, "Enhancing performance of network-on-chip architectures with millimeter-wave wireless interconnects," *Application-specific Systems Architectures and Processors (ASAP), 21st IEEE International Conference on*, pp. 73–0, 2010.
- [20] D. Zhao, Y. Wang, and S. Member, "SD-MAC : Design and Synthesis of a MAC Protocol for Wireless Network-on-Chip," *Computers, IEEE Transactions on*, vol. 57, no. 9, pp. 1230–1245, 2008.
- [21] A. Ganguly, K. Chang, S. Deb, P. P. Pande, B. Belzer, and C. Teuscher, "Scalable Hybrid Wireless Network-on-Chip Architectures for Multi-Core Systems," *IEEE Transactions on Computers*, vol. 60, no. 10, pp. 1485–1502, 2010.
- [22] I. Llatser, C. Kremers, D. Chigrin, J. M. Jornet, M. C. Lemme, A. Cabellos-Aparicio, and E. Alarcón, "Radiation Characteristics of Tunable Graphennas in the Terahertz Band," *Radioengineering Journal*, vol. 21, no. 4, 2012.
- [23] M. Tamagnone, J. S. Goñamez-Diñáñez, J. R. Mosig, and J. Perruisseau-Carrier, "Analysis and design of terahertz antennas based on plasmonic resonant graphene sheets," *Journal of Applied Physics*, vol. 112, p. 114915, 2012.
- [24] J. Jornet and I. Akyildiz, "Channel modeling and capacity analysis for electromagnetic wireless nanonetworks in the terahertz band," *Wireless Communications, IEEE Transactions on*, vol. 10, no. 10, pp. 3211–3221, 2011.
- [25] E. Gul, B. Atakan, and O. Akan, "Nanons: a nanoscale network simulator framework for molecular communications," *Nano Communication Networks*, vol. 1, pp. 138–156, Oct. 2011.
- [26] I. Llatser, I. Pascual, N. Garralda, A. Cabellos-Aparicio, and E. Alarcon, "N3sim: a

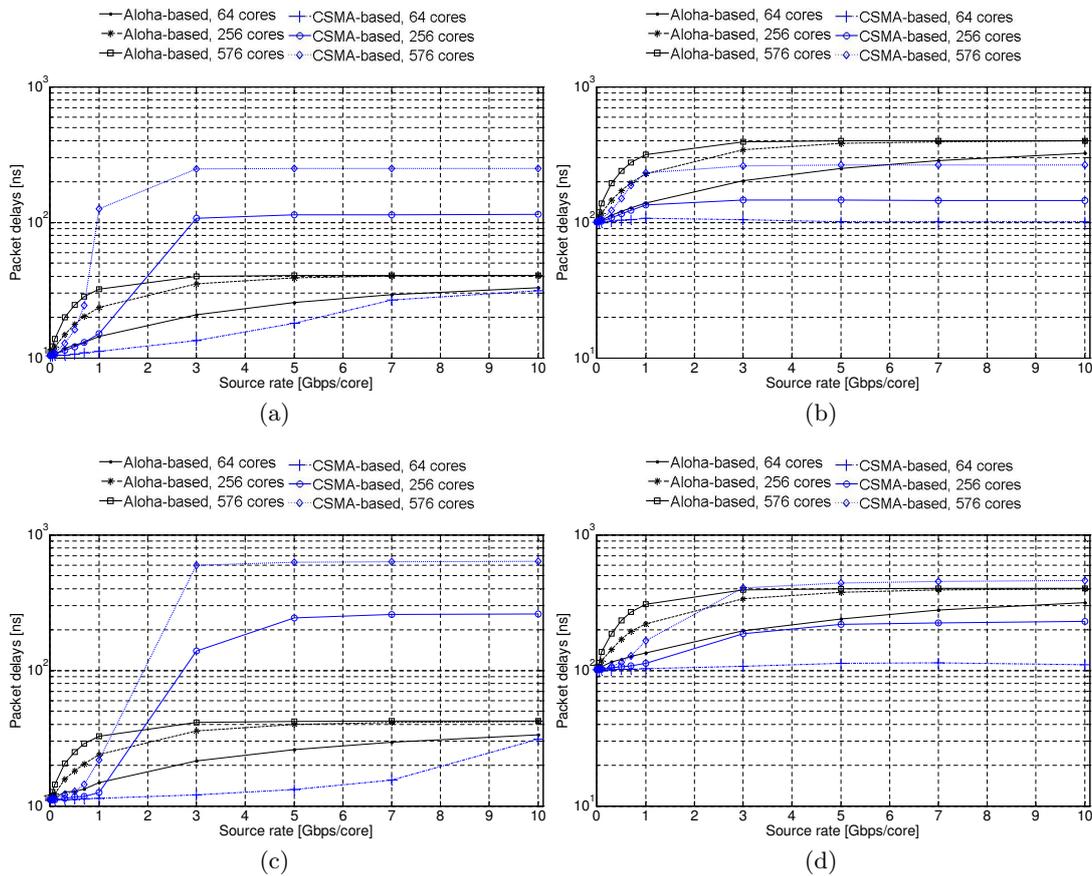


Figure 5: Packet delay, evaluated when (a) packet size = 312 bits and $W_D = 10$ ns, (b) packet size = 312 bits and $W_D = 100$ ns, (c) packet size = 1024 bits and $W_D = 10$ ns, and (d) packet size = 1024 bits and $W_D = 100$ ns.

- simulation framework for diffusion-based molecular communication,” *IEEE Technical Committee on Simulation*, vol. 8, pp. 3–4, 2011.
- [27] L. Felicetti, M. Femminella, and G. Reali, “A simulation tool for nanoscale biological networks,” *Nano Communication Networks*, Oct. 2011.
- [28] G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, “Nano-sim: simulating electromagnetic-based nanonetworks in the network simulator 3,” in *Proc. of Workshop on NS-3 (held in conjunction with SIMUTools 2013)*, (Cannes, France), Mar. 2013.
- [29] G. Piro, L. A. Grieco, G. Boggia, and P. Camarda, “Simulating wireless nano sensor networks in the ns-3 platform,” in *Proc. of Workshop on Performance Analysis and Enhancement of Wireless Networks, PAEWN*, (Barcelona, Spain), Mar. 2013.
- [30] IEEE, “P1906.1 - recommended practice for nanoscale and molecular communication framework, <http://standards.ieee.org/develop/project/1906.1.html> [accessed: 9/10/2014].”
- [31] H. C. Freitas and P. O. A. Navaux, “Evaluating On-Chip Interconnection Architectures for Parallel Processing,” in *The 11th IEEE International Conference on Computational Science and Engineering - Workshops*, pp. 188 – 193, 2008.
- [32] S. Ma, N. Jerger, and Z. Wang, “Supporting efficient collective communication in NoCs,” in *Proceedings of the 2012 IEEE 18th International Symposium on High-Performance Computer Architecture*, pp. 1–12, 2012.
- [33] D. Wentzlaff, P. Griffin, H. Hoffmann, L. Bao, B. Edwards, C. Ramey, M. Mattina, C.-C. Miao, J. F. B. III, and A. Agarwal, “On-chip interconnection architecture of the tile processor,” *IEEE Micro*, vol. 27, no. 5, pp. 15–31, 2007.
- [34] W. Huang, K. Rajamani, M. Stan, and K. Skadron, “Scaling with design constraints: Predicting the future of big chips,” *IEEE Micro*, pp. 16–29, 2011.
- [35] J. Chan, G. Hendry, A. Biberman, K. Bergman, and L. P. Carloni, “PhoenixSim: A Simulator for Physical-Layer Analysis of Chip-Scale Photonic Interconnection Networks,” in *Proceedings of the Conference on Design, Automation and Test in Europe (DATE)*, pp. 691–696, 2010.